

WHAT IS CLAIMED IS:

1 1. A method of selecting a set of tag nucleic acids with minimal cross
2 hybridization to a nucleic acid, said method comprising providing a list of tag nucleic
3 acids, and excluding nucleic acids from the list of tag nucleic acids which cross hybridize
4 to a single complementary nucleic acid under stringent conditions, thereby providing a set
5 of selected tag nucleic acids with minimal cross hybridization to the nucleic acid.

1 2. The method of claim 1, wherein the method of selecting tag nucleic
2 acids further comprises:

3 selecting a first tag nucleic acid from the list of tag nucleic acids;
4 selecting a second tag nucleic acid from the list of tag nucleic acids;
5 comparing the sequence of the first tag nucleic acid to the sequence of the second
6 tag nucleic acid; and,
7 determining that the second tag nucleic acid hybridizes to the complement of the
8 first tag nucleic acid with a selected thermal binding stability, thereby excluding the
9 second nucleic acid from the selected set of nucleic acid tags.

1 3. The method of claim 2, wherein the method comprises rejecting or
2 selecting each tag in the list of tags in order.

1 4. The method of claim 2, wherein tags are not selected if they have
2 more than 8 contiguous nucleotides in common with any previous tag.

1 5. The method of claim 2, wherein the method comprises rejecting or
2 selecting tags in complementary pairs, wherein each selected tag has a complementary
3 selected tag.

1 6. The method of claim 1, wherein the method further comprises
2 selecting a thermal binding stability for the tags, and excluding all tag nucleic acids from
3 the list of tag nucleic acids which do not have the selected thermal binding stability.

1 7. The method of claim 6, wherein the thermal binding stability is
2 selected by specifying a ratio of G+C to A+T nucleotides for the tag nucleic acids, and
3 specifying a length for the tag nucleic acids.

1 8. The method of claim 1, wherein the method further comprises
2 excluding tags which contain self-complementary regions from the list of tags.

1 9. The method of claim 6, wherein the regions of self complementarity
2 are greater than 4 nucleotides in length.

1 10. The method of claim 1, wherein the tags are between 15 and 30
2 nucleotides in length.

3 11. The method of claim 1, wherein the tags are between 10 and 100
4 nucleotides in length.

1 12. The method of claim 1, wherein the tags are 20 nucleotides in
2 length.

1 13. The method of claim 1, wherein the method further comprises
2 selecting a complementary probe nucleic acid for tags in the selected tag set, wherein
3 each tag sequence is complementary to one probe sequence, and the thermal binding
4 stability between each tag and each complementary probe is substantially similar.

1 14. The method of claim 13, wherein all of the tags have the same
2 length and the same GC to AT ratio.

1 15. The method of claim 1, wherein the method further comprises
2 selecting a constant region subsequence shared by all tag nucleic acids, thereby
3 determining the nucleotide position of variable nucleotides in the tags.

1 16. The method of claim 15, wherein the method further comprises
2 providing a set of probe nucleic acids by determining the complement to each variable
3 nucleotide in each tag nucleic acid, and selecting a probe comprising a corresponding
4 complementary nucleotide for each nucleotide in the variable tag sequence, which probe
5 does not hybridize to the constant region of the tag nucleic acid, thereby providing a
6 selected set of probes.

1 17. The method of claim 1, wherein the method further comprises
2 removing tag nucleic acids which have fewer than two nucleotide differences when
3 aligned for maximal sequence correspondence.

4 18. The method of claim 17, wherein:
5 the total number of nucleotides in each of the selected sets is identical;
6 the number of G+C nucleotides in each tag in the selected set is identical; and,
7 the overall number of A+G nucleotides in each of the variable regions of the tags
8 is even.

1 19. The method of claim 1, wherein the method further comprises
2 removing tag nucleic acids which have fewer than 5 nucleotide differences when aligned
3 for maximal sequence correspondence.

1 20. The method of claim 1, wherein tags which contain 4 contiguous
2 nucleotides selected from the group consisting of 4 X residues, 4 Y residues and 4 Z
3 residues, are eliminated from the tag set, wherein X is selected from the group consisting
4 of G and C, Y is selected from the group consisting of G and A, and Z is selected from
5 the group consisting of A and T.

1 21. A composition comprising a set of tag nucleic acids, which set of
2 tag nucleic acids comprises a plurality of tag nucleic acids, which tag nucleic acids
3 comprise a variable region;

4 which variable region for each tag nucleic acid in the set of tag nucleic acids has
5 the same T_m , the same G+C to A+T ratio, the same length and does not cross-hybridize
6 to a probe nucleic acid; and,

7 wherein the tag nucleic acids in the set of tag nucleic acids cannot be aligned with
8 less than two differences between any two of the tag nucleic acids in the set of tag nucleic
9 acids.

1 22. The composition of claim 21, wherein the tags comprise a constant
2 region.

1 23. The composition of claim 21, wherein the variable region of each of
2 the tag nucleic acids in the tag set comprises less than two C nucleotides.

1 24. The composition of set of claim 21, wherein the variable region of
2 the tag nucleic acids from the set of tag nucleic acids comprises an even number of A+G
3 nucleotides.

1 25. A method of labeling a composition, comprising associating a tag
2 nucleic acid with the composition, wherein the tag nucleic acid is selected from a group
3 of tag nucleic acids which do not cross-hybridize and which have a substantially similar
4 T_m .

1 26. The method of claim 25, further comprising detection of the tag
2 nucleic acid.

1 27. The method of claim 25, further comprising detection of the tag
2 nucleic acid by labeling the nucleic acid and hybridizing the nucleic acid to a solid
3 substrate, which substrate comprises an array of probe nucleic acids selected to hybridize
4 to the group of tag nucleic acids.

1 28. The method of claim 25, further comprising amplification of the tag
2 nucleic acids, thereby providing amplified tag nucleic acids and detection of the amplified

3 tag nucleic acids by hybridization to an array of probes complementary to the tag nucleic
4 acids.

1 29. The method of claim 28, wherein the tag nucleic acids are amplified
2 using the polymerase chain reaction.

1 30. The method of claim 28, wherein the amplified tag nucleic acids are
2 labeled with a fluorescent label.

1 31. A method of pre-selecting experimental probes in an oligonucleotide
2 probe array, wherein the probes have substantially uniform hybridization properties and
3 do not cross hybridize, comprising:

4 selecting a ratio of G+C to A+T nucleotides shared by the experimental probes in
5 the array;

6 determining all possible 4 nucleotide subsequences for variable nucleic acids in the
7 probes of the array; and

8 excluding all probes from the array which contain prohibited 4 nucleotide sub-
9 sequences; wherein 4 nucleotide subsequences are prohibited when the nucleotide
10 subsequences are selected from the group consisting of self-complementary probes, A₄
11 probes, T₄ probes, [G,C]₄ probes, and probes complementary to constant region sub-
12 sequences.

1 32. The method of claim 31, wherein the method further comprises
2 selecting a length for the probes in the array, thereby providing selected length
3 probes;

4 selecting a constant region subsequence shared by all selected length probes in the
5 array, thereby determining the nucleotide position of variable nucleic acids in the probes
6 of the array; and

7 providing that the overall number of A+G nucleotides in the probes of the array is
8 even.

1 33. The method of claim 31, wherein the method further comprises
2 selecting control probes for addition to the array.

1 34. A method of detecting a plurality of nucleic acids in a sample,
2 comprising

3 (i) providing an array of experimental oligonucleotide probes, which probes do not
4 cross hybridize under stringent conditions, wherein the ratio of G+C bases in each probe
5 is substantially identical;

6 wherein the probes of the array are arranged into probe sets in which each
7 probe set comprises a homogeneous population of oligonucleotide probes;

8 (ii) hybridizing said array of oligonucleotides to the sample under stringent
9 hybridization conditions; and

10 (iii) detecting hybridization of the nucleic acids to the array of oligonucleotide
11 probes.

1 35. The method of claim 34, wherein the probes of the array
2 specifically hybridize to at least one nucleic acid in the sample.

1 36. The method of claim 34, wherein the nucleic acids comprise tag
2 sequences, which tag sequences bind to the probes of the array.

1 37. An array of oligonucleotide probes comprising a plurality of
2 experimental oligonucleotide probe sets attached to a solid substrate, wherein
3 each experimental oligonucleotide probe set in the array hybridizes to a different
4 target nucleic acid under stringent hybridization conditions;
5 each oligonucleotide probe in the probe sets of the array comprises a variable
6 region; and wherein
7 the nucleic acid probes do not cross-hybridize in the array.

1 38. The array of claim 37, wherein each probe set in the array a
2 constant region, wherein the variable region does not cross hybridize with the constant
3 region under stringent hybridization conditions.

1 39. The array of claim 37, wherein each probe set in the array differs
2 from every other probe set in the array by the arrangement of at least two nucleotides in
3 the probes of the probe set.

1 40. The array of claim 37, wherein the ratio of G+C bases in each
2 probe for each experimental probe set is substantially identical.

1 41. The array of claim 37, wherein the array comprises a plurality of
2 probe sets selected from the output group of probes produced by running tags.ccp.

1 42. The array of claim 37, wherein the array further comprises a
2 nucleic acid bound to a probe in the array.

1 43. The array of claim 37, wherein the array further comprises control
2 probes.

1 44. A method of detecting a target nucleic acid comprising providing a
2 population of nucleic acids to an array of oligonucleotide probes and monitoring
3 hybridization of the test nucleic acids to the probes in the array, wherein the array of
4 oligonucleotide probes comprises a plurality of experimental oligonucleotide probe sets
5 attached to a solid substrate, wherein

6 each experimental oligonucleotide probe set in the array hybridizes to a different
7 target nucleic acid under stringent hybridization conditions;

8 each oligonucleotide probe in the probe sets of the array comprises variable
9 region; and wherein

10 the nucleic acid probes do not cross-hybridize in the array.

1 45. The method of claim 44, wherein the probes of the array comprise a
2 constant region, wherein the variable region does not cross hybridize with the constant
3 region under stringent hybridization conditions.

1 46. The method of claim 44, wherein the array comprises a control
2 probe, and wherein the method further comprises hybridizing a nucleic acid
3 complementary to the control probe to the array.

1 47. A plurality of recombinant cells comprising tag nucleic acids
2 selected from a set of tag nucleic acids, which set of tag nucleic acids comprises a
3 plurality of tag nucleic acids, which tag nucleic acids comprise a variable region;
4 which variable region for each tag nucleic acid in the set of tag nucleic acids has
5 the same T_m , the same G+C to A+T ratio, the same length and does not cross-hybridize;
6 and,
7 wherein the tag nucleic acids in the set of tag nucleic acids cannot be aligned with
8 less than two differences between any two of the tag nucleic acids in the set of tag nucleic
9 acids.

1 48. The recombinant cell of claim 47, wherein the tags further comprise
2 a constant region, wherein the variable region does not cross hybridize with the constant
3 region under stringent hybridization conditions.

1 49. The recombinant cell of claim 47, wherein the cell is selected from
2 a library of genetically distinct recombinant cells.

1 50. The recombinant cell of claim 47, wherein the cell is a eukaryotic
2 cell.

1 51. The recombinant cell of claim 47, wherein the cell is a prokaryotic
2 cell.

1 52. The recombinant cell of claim 47, wherein the cell is a yeast cell.

1 53. A kit comprising an array of oligonucleotides, wherein
2 the array of oligonucleotide probes comprises a plurality of experimental
3 oligonucleotide probe sets attached to a solid substrate;

4 each experimental oligonucleotide probe set in the array hybridizes to a different
5 target nucleic acid under stringent hybridization conditions;

6 each oligonucleotide probe in the probe sets of the array comprises a variable
7 region; and

8 the nucleic acid probes do not cross-hybridize in the array.

1 54. The kit of claim 53, wherein each oligonucleotide in the array
2 further comprises a constant region, wherein the variable region does not cross hybridize
3 with the constant region under stringent hybridization conditions.

1 55. The kit of claim 53, wherein the kit further comprises a plurality of
2 tag nucleic acids complementary to the experimental oligonucleotide probes in the array.

1 56. The kit of claim 53, wherein the array further comprises control
2 oligonucleotide probes.

1 57. The kit of claim 53, wherein the kit further comprises PCR
2 reagents, a container and instructions.